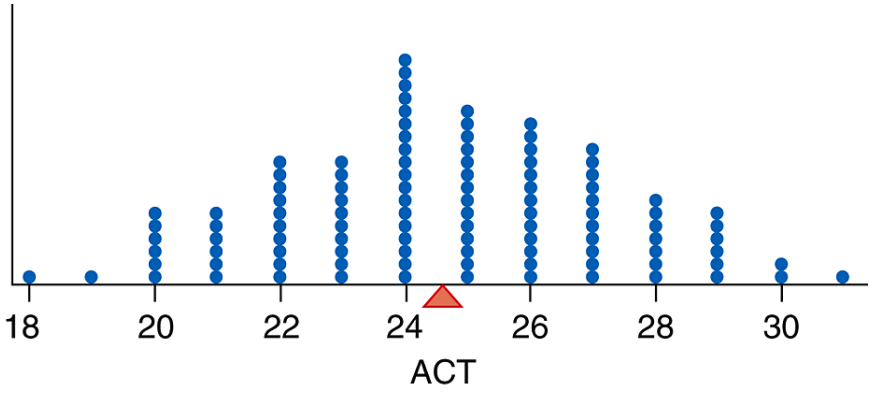
3.1: Summaries for Symmetric Distributions

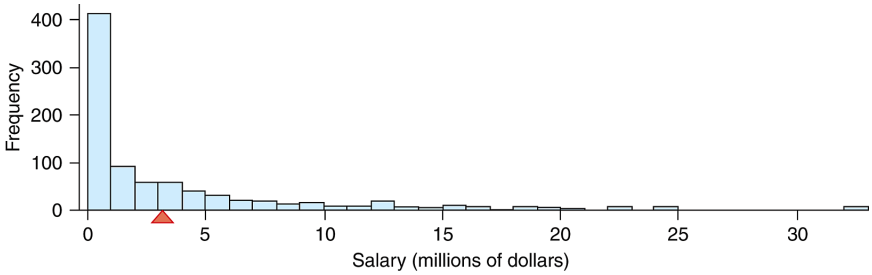
Definition.
Given a collection of data $\{x_1, x_2, \ldots, x_n\}$, the **mean** of the data is the arithmetic mean:

$$\bar{x} = \frac{x_1}{n} + \frac{x_2}{n} + \cdots + \frac{x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Example. An instructor at Peoria Junior College in Illinois collected data from two classes, including the students' ACT scores. Below is the distribution of self-reported ACT scores for one statistics class:



Example. The winnings of the top-ranked professional tennis players in the 2018 season are given in the graph below:

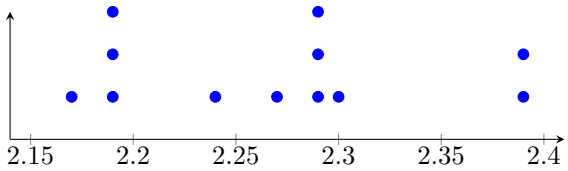


Note:

- When the distribution is roughly symmetric, the mean represents a typical value in the data.
- The mean is *not* a good estimate of a typical value of a skewed distribution.

Example. According to GasBuddy.com (a website that invites people to submit prices at local gas stations), the prices of 1 gallon of regular gas at 12 service stations near the downtown area of Austin, TX, were as follows one winter day in 2018:

\$2.19	\$2.19	\$2.39	\$2.19
\$2.24	\$2.39	\$2.27	\$2.29
\$2.17	\$2.29	\$2.30	\$2.29

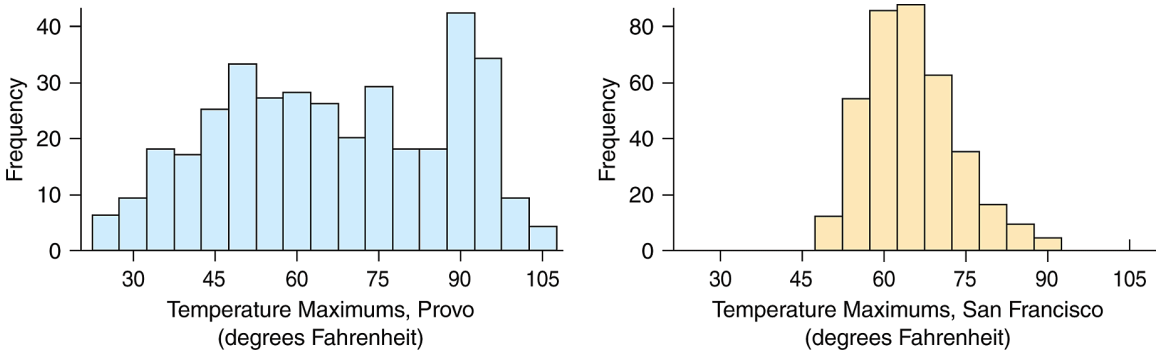


Find the mean price of a gallon of regular gas at these service stations, and interpret the result.

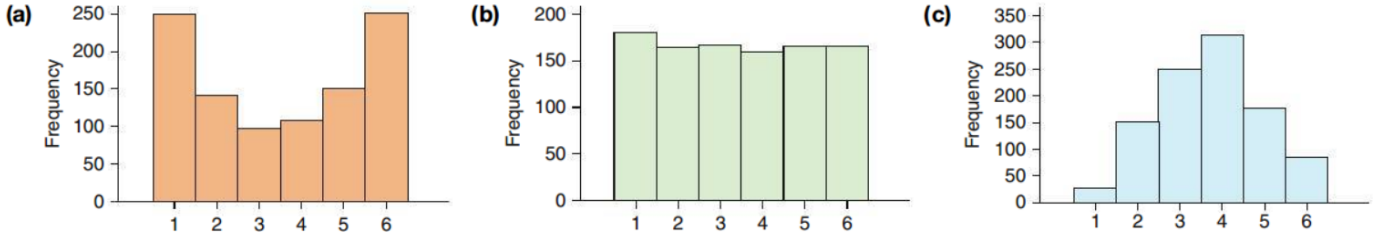
Definition.
The **standard deviation** is a number that measures how far the typical observation is from the mean. For symmetric, unimodal distributions, a majority of the data is within one standard deviation of the mean. The standard deviation is given by

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

Example. The histograms below show daily high temperatures in degrees Fahrenheit recorded over one year in Provo, Utah (left), and San Francisco, California (right). Which city do we expect to have a higher standard deviation?



Example. Below are three histograms representing distributions with the same mean. Which distribution has the largest standard deviation? Which has the smallest?



Example. Recall the data set of gas prices from before. Use StatCrunch to compute the standard deviation of this data set and interpret the result: statcrunch.com

\$2.19	\$2.19	\$2.39	\$2.19
\$2.24	\$2.39	\$2.27	\$2.29
\$2.17	\$2.29	\$2.30	\$2.29

1. Click "Open StatCrunch"
2. Enter data into spreadsheet
3. Under the "Stat" menu, select "Summary Stats" then "Columns" or "Rows"

Definition.
The **variance** is the standard deviation squared:

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$